Merida virus, a putative novel rhabdovirus discovered in *Culex* and *Ochlerotatus* spp. mosquitoes in the Yucatan Peninsula of Mexico

Jermilia Charles, Andrew E. Firth, Maria A. Loroño-Pino, Julian E. García-Rejon, Jose A. Farfan-Ale, W. Ian Lipkin, Bradley J. Blitvich and Thomas Briese

Correspondence
Bradley J. Blitvich
blitvich@iastate.edu

Departments of veterinary microbiology and preventive medicine, college of veterinary medicine, Iowa State University, Ames, IA, USA

Department of pathology, University of Cambridge, Cambridge, UK

Laboratorio de Arbovirología, Centro de Investigaciones Regionales Dr Hideyo Noguchi, Universidad Autónoma de Yucatán, Mérida, Yucatán, México

Center for Infection and Immunity, Mailman School of Public Health, Columbia University, New York, NY, USA

Sequences corresponding to a putative, novel rhabdovirus [designated Merida virus (MERDV)] were initially detected in a pool of *Culex quinquefasciatus* collected in the Yucatan Peninsula of Mexico. The entire genome was sequenced, revealing 11,798 nt and five major ORFs, which encode the nucleoprotein (N), phosphoprotein (P), matrix protein (M), glycoprotein (G) and RNA-dependent RNA polymerase (L). The deduced amino acid sequences of the N, G and L proteins have no more than 24, 38 and 43% identity, respectively, to the corresponding sequences of all other known rhabdoviruses, whereas those of the P and M proteins have no significant identity with any sequences in GenBank and their identity is only suggested based on their genome position. Using specific reverse transcription-PCR assays established from the genome sequence, 27,571 *C. quinquefasciatus* which had been sorted in 728 pools were screened to assess the prevalence of MERDV in nature and 25 pools were found positive. The minimal infection rate (calculated as the number of positive mosquito pools per 1000 mosquitoes tested) was 0.9, and similar for both females and males. Screening another 140 pools of 5484 mosquitoes belonging to four other genera identified positive pools of *Ochlerotatus* spp. mosquitoes, indicating that the host range is not restricted to *C. quinquefasciatus*. Attempts to isolate MERDV in C6/36 and Vero cells were unsuccessful. In summary, we provide evidence that a previously undescribed rhabdovirus occurs in mosquitoes in Mexico.

INTRODUCTION

The family Rhabdoviridae (order Mononegavirales) is composed of a large and versatile group of viruses that are ubiquitous in nature (Kuzmin et al., 2009). The family consists of 11 genera as well as several viruses that have not yet been assigned to a genus (<http://www.ictvonline.org/virusTaxonomy.asp>). Virions have a distinctive bullet or cone-shaped morphology or appear bacilli-form. Rhabdoviruses have broad host ranges that include humans and other terrestrial mammals, birds, reptiles, fish, insects and plants (Hoffmann et al., 2005; Kuzmin et al., 2009; Mann & Dietzgen, 2014). Many rhabdoviruses are transmitted to vertebrate and plant hosts by insect vectors in which they replicate (Hogenhout et al., 2003; Ammar et al., 2009). Rhabdoviruses have a negative-sense ssRNA genome of ~11–16 kb. A universal feature of the rhabdovirus genome is the presence of at least five genes that code in 3'→5' order for the structural proteins: nucleoprotein (N), phosphoprotein (P), matrix protein (M), glycoprotein (G) and RNA-dependent RNA polymerase (L) (Fu, 2005; Walker et al., 2015). Each ORF is flanked by conserved cis-acting transcription initiation and transcription
termination/polyadenylation signals that regulate mRNA expression. Rhabdovirus genomes are often interspersed with smaller ORFs that encode accessory proteins, most of which are of unknown function (Walker et al., 2011). ORFs that encode accessory proteins can occur as alternative or overlapping ORFs within the major structural protein genes or as independent ORFs in the intergenic regions that separate the structural protein genes. Leader and trailer sequences are located at the 3′ and 5′ termini of the rhabdovirus genome, respectively. These sequences are non-coding, A/U-rich and usually 50–100 nt in length. The first 10–20 nt of the leader and trailer sequences commonly exhibit partial complementarity, and function as promoter sequences required for the initiation of genome and anti-genome replication, respectively (Fu, 2005; Walker et al., 2015).

Many novel rhabdoviruses have been discovered in recent years due to the advent of unbiased high-throughput sequencing (UHTS) (Quan et al., 2010; Ito et al., 2013; Kading et al., 2013; Tokarz et al., 2014; Binger et al., 2015; Sakai et al., 2015; Stremlau et al., 2015), including Bas-Congo virus which was associated with an outbreak of acute haemorrhagic fever in humans in the Democratic Republic of Congo (Grard et al., 2012). Here, we report the genomic organization and prevalence of an apparently novel rhabdovirus tentatively named Merida virus (MERDV) that was discovered by UHTS in mosquitoes in the Yucatan Peninsula of Mexico.

RESULTS

Discovery of MERDV

UHTS of total RNA from a pool of Culex quinquefasciatus collected in Merida in the Yucatan Peninsula of Mexico generated a ~11 kb sequence corresponding to a putative, novel rhabdovirus tentatively named MERDV. Several other novel virus-like sequences were also discovered and the data will be presented elsewhere. The MERDV genome terminal sequences were identified using a combination of 5′ and 3′ RACE and Sanger sequencing. Difficulties were encountered during the 3′ RACE because the reverse primer bound preferentially to an A-rich region located slightly upstream of the 3′ terminus. The 3′ end of the genome was eventually identified by taking advantage of the partial complementarity that exists between the 5′ and 3′ termini of the rhabdovirus genome. The 3′ end of the MERDV genome was amplified and sequenced using a reverse primer designed from the inverse complement of the 24 nt sequence at the distal end of the 5′ terminus. Therefore, our sequence may contain nucleotide errors in the 24 nt primer-binding region at the ultimate 3′ end of the genome.

Genomic organization

The MERDV genome consists of 11 798 nt (GenBank accession number KU194360) and its organization is consistent with that of the classical rhabdovirus genome: short leader and trailer sequences (68 and 74 nt, respectively) flank five structural protein genes in the order 3′-N–P–M–G–L-5′ (Fig. 1). The terminal nucleotides at the 5′ end of the MERDV genome are 5′-ACG-3′ and these same trinucleotides are located at the 5′ termini of other related rhabdovirus genomes (Gubala et al., 2008, 2011; Kuwata et al., 2011; Zhu et al., 2011). The leader and trailer sequences are 59 and 58 % A/U-rich, respectively. Each ORF is separated by a non-coding region of 54–153 nt that contains transcription initiation and termination sequences identified as AACAU and CAUG[A]7, respectively, through sequence alignment of conserved nucleotides. The only exception to these consensus motifs is the CUUG[A]7 transcription termination sequence that regulates M mRNA expression (Table 1).

The genomic location and length of each predicted ORF is shown in Table 2. The ORF that encodes the N protein consists of 1437 nt. The predicted translation product is most closely related to the corresponding protein of Culex tritaeniorhynchus rhabdovirus (CTRV; 24 % identity and 45 % similarity) which was recently discovered in C. tritaeniorhynchus in Japan (Kuwata et al., 2011) and Yongia tick virus 2 (also 24 % identity and 45 % similarity) from Haemaphysalis hystricis ticks in China (Li et al., 2015). The next two ORFs encode translation products that have no significant identity with any other sequences in GenBank. These two ORFs are assumed to encode the P and M proteins based on their positions in the genome. The ORF that encodes the G protein consists

---

**Fig. 1.** Coding capacity of MERDV sequence. Main ORFs, as well as minor ORFs unlikely to be expressed, are indicated along a scaled representation of the anti-genomic strand. The size of minor ORFs is indicated by their codon number: blue, frame 1; green, frame 2; red, frame 3.
of 1530 nt and the predicted translation product is most closely related to the corresponding protein of CTRV (38 % identity and 58 % similarity). The next closest match is to a tandem rhabdovirus-like glycoprotein domain repeat sequence identified by the *Aedes aegypti* sequencing consortium (Nene et al., 2007). The largest ORF in the MERDV genome consists of 6411 nt and encodes the L protein. The predicted translation product is also most closely related to the corresponding region of CTRV (44 % identity and 65 % similarity). Minor ORFs of 102, 62, 83 and 75 codons overlap the N, P, L and L genes, respectively, but are not obviously accessible via ribosomal scanning (Fig. 1). The AUG of a fifth ORF (75 codons) that overlaps the 5' end of the P ORF is apparently upstream of the P mRNA transcription start site, so is also unlikely to be accessible for translation.

**Predicted domains and post-translational modifications**

The G proteins of rhabdoviruses have several common characteristics, including the presence of two to six potential N-linked glycosylation sites, 12 well-conserved cysteine residues, an N-terminal signal peptide, a transmembrane domain and a C-terminal short hydrophilic cytoplasmic domain (Coll, 1995; Walker & Kongsuwan, 1999). The G protein of MERDV is predicted to contain four potential N-linked glycosylation sites (one less than the G protein of CTRV) (Kuwata et al., 2011). All 12 conserved cysteines are present with two additional cysteines located at residues 12 and 485 (in CTRV one additional cysteine is present) (Kuwata et al., 2011). The G protein of MERDV is predicted to contain a signal peptide at residues 1–17, a

### Table 1. Non-coding sequences in the MERDV genome

<table>
<thead>
<tr>
<th>Gene</th>
<th>Non-coding sequences at the start of the gene</th>
<th>Non-coding sequences at the end of the gene</th>
<th>Intergenic region*</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>ACGAAAAACAAAAAATCCCCCCAATCACAGTCAGA</td>
<td>GCCCUCACCUCGGAGCCUCCGGAUGCCCA</td>
<td>ACUCC</td>
</tr>
<tr>
<td></td>
<td>ATCCCATTTTGTTCTGGATGTGAAT</td>
<td>CAGGUUAACUGAAAAA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CTTAACAUAACGUAAAUAAUAAACUCUA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AUUAUGACACUUAUUGGAAUUCUUCUGGAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GCGCGAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>AACAUAAACUACUCGGACUGAACCCGAUC</td>
<td>CAUGAAAAAAA</td>
<td>CUCC</td>
</tr>
<tr>
<td></td>
<td>AUUCACA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>AACAUCACUCACUGAGACUAUCCCGAGU</td>
<td>AUCCCCCUUUAUAGACUUGGACCUUGUAAUU</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UAAUUCUGGCAAC</td>
<td>CCAAAUAAAAAGACUAUACUAACUAGAAAUCU</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>AACAUAACUAGAACAACUGAAGAAAAGUUGCCUC</td>
<td>AGACCGUAUGAUAUCCGCUGUUCUGUUGUUU</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UAGGAUACUGAGGAAAGCGGGGGAAGAAAAGCAAUGAAGAGAGA</td>
<td>GUCUGAGCAGUAAUAGAUGGACCUUGCUCAGCUACU</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>AACGAUAACACACAGAGGAAAGAAGACACAGAGAGAGAGAGAGA</td>
<td>AUGUGCUAAGCUUAAAUAAACUGACGUAUGGGAUA</td>
<td></td>
</tr>
</tbody>
</table>

*Intergenic region located immediately downstream of the gene of interest that is neither translated nor transcribed to mRNA.

†No intergenic region present due to gene overlap by 25 nt between the M and G genes and 13 nt between the G and L genes.

### Table 2. Predicted locations and lengths of the ORFs encoded by the MERDV genome

<table>
<thead>
<tr>
<th>Protein</th>
<th>ORF genomic location</th>
<th>ORF length (nt)</th>
<th>Protein length (aa)</th>
<th>Protein mass (kDa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>131–1567</td>
<td>1437</td>
<td>478</td>
<td>54.2</td>
</tr>
<tr>
<td>P</td>
<td>1660–2862</td>
<td>1203</td>
<td>400</td>
<td>43.8</td>
</tr>
<tr>
<td>M</td>
<td>2917–3477</td>
<td>564</td>
<td>187</td>
<td>21.0</td>
</tr>
<tr>
<td>G</td>
<td>3586–5112</td>
<td>1530</td>
<td>509</td>
<td>56.8</td>
</tr>
<tr>
<td>L</td>
<td>5266–11676</td>
<td>6411</td>
<td>2136</td>
<td>241.2</td>
</tr>
</tbody>
</table>
tested by RT-PCR using MERDV-specific primers RHAB-

Another 5484 mosquitoes belonging to seven species were tested by RT-PCR using MERDV-specific primer pairs, whilst all other species were negative. The MIRs for MERDV in mosquitoes collected in both study areas and during both time periods. See Tables 3 and 4.

Detection of MERDV sequence in other mosquito species

A total of 27 571 C. quinquefasciatus, sorted into 728 pools of up to 50 individuals, were screened by MERDV-specific reverse transcription (RT)-PCR. Collections were made using mosquito magnets at five study sites (Cozumel Island, Merida, Sian Ka’an, Tixkokob and Tzucacab) in 2007–2008. Mosquitoes had been sorted into 140 pools (20 pools per species) and all were female. MERDV RNA was detected in three pools of O. taeniorynchus and in three pools of O. trivittatus using both primer pairs, whilst all other species were negative. The MERDV MIRs in O. taeniorynchus and O. trivittatus were calculated as 3.0 and 4.1, respectively. All six PCR products generated using primers RHAB-121-for and RHAB-280-rev were analysed by Sanger sequencing. The resulting 114 nt sequences had at least 99.1 % nucleotide identity with the corresponding region of the MERDV genome sequence identified in C. quinquefasciatus (data not shown).

Prevalence in C. quinquefasciatus

A total of 27 571 C. quinquefasciatus, sorted into 728 pools of up to 50 individuals, were screened by MERDV-specific reverse transcription (RT)-PCR. Collections were made in Merida and Tixkokob in 2007–2008 using mosquito magnets, and in Merida in 2013 using Centers for Disease Control and Prevention (CDC) backpack-mounted aspirators. Mosquitoes obtained in 2007–2008 were tested according to gender, whereas those from 2013 were not. Overall, 256 pools (8038 mosquitoes) were composed of females, 195 pools (7196 mosquitoes) were composed of males and 277 pools (12 337 mosquitoes) were of mixed gender. Twenty-five mosquito pools were positive for MERDV RNA and the overall minimal infection rate (MIR; calculated as the number of positive mosquito pools per 1000 mosquitoes tested) was 0.9. The MIRs for female and male mosquitoes were similar (1.1 and 1.0, respectively). Evidence of MERDV infection was detected in mosquitoes collected in both study areas and during both time periods. See Tables 3 and 4.

Detection of MERDV sequence in other mosquito species

Another 5484 mosquitoes belonging to seven species were tested by RT-PCR using MERDV-specific primers RHAB-
Discovery of a novel rhabdovirus in Mexican mosquitoes

**Table 3.** MIRs for MERDV in *C. quinquefasciatus* in the Yucatan Peninsula of Mexico, 2007–2008 and 2013

<table>
<thead>
<tr>
<th>Study site</th>
<th>Date</th>
<th>No. mosquitoes tested</th>
<th>No. pools tested (positive)</th>
<th>MIR*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tzikkokob†</td>
<td>2007–2008</td>
<td>9071</td>
<td>247 (2)</td>
<td>0.2</td>
</tr>
<tr>
<td>Merida†</td>
<td>2007–2008</td>
<td>6163</td>
<td>204 (14)</td>
<td>2.3</td>
</tr>
<tr>
<td>Merida‡</td>
<td>2013</td>
<td>12 337</td>
<td>277 (9)</td>
<td>0.7</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>27 571</td>
<td>728 (25)</td>
<td>0.9</td>
</tr>
</tbody>
</table>

*Expressed as the number of positive mosquito pools per 1000 mosquitoes tested.
†Mosquitoes were collected outdoors using mosquito magnets.
‡Mosquitoes were collected inside private residences using CDC backpack-mounted aspirators.

---

**Fig. 2.** Phylogenetic tree for MERDV and selected other rhabdovirus sequences. L protein amino acid sequences were aligned using MUSCLE (Edgar, 2004). A maximum-likelihood phylogenetic tree was estimated using the Bayesian Markov chain Monte Carlo method implemented in MrBayes version 3.2.3 (Ronquist et al., 2012) sampling across the default set of fixed amino acid rate matrices with 10 million generations, discarding the first 25 % as burn-in. The original figure was produced using FigTree (http://tree.bio.ed.ac.uk/software/figtree/). The tree is midpoint-rooted and selected nodes are labelled with posterior probability values. Rhabdovirus genera, where defined, are labelled on the far right. GenBank accession numbers are indicated next to virus names. Bar indicates amino acid substitutions per site.

---

http://jgv.microbiologyresearch.org
DISCUSSION

The advent of UHTS has resulted in the discovery of many novel rhabdoviruses (Grard et al., 2012; Ito et al., 2013; Kading et al., 2013; Tokarz et al., 2014; Binger et al., 2015; Sakai et al., 2015; Stremlau et al., 2015), including several which were isolated from Anopheles, Culex, Ochlerotatus and Psorophora spp. mosquitoes (Quan et al., 2010; Coffey et al., 2014; Vasilakis et al., 2014). Here, we report the discovery of a putative, novel rhabdovirus in Culex and Ochlerotatus spp. mosquitoes from the Yucatan Peninsula of Mexico.

It is hypothesized that the majority of Rhabdoviruses are hosted by insects (Li et al., 2015), and numerous cyto-, nucleo- and dimarhabdoviruses are transmitted by arthropods to their plant or vertebrate hosts (Hogenhout et al., 2003; Bourhy et al., 2005). This includes the bite of haematophagous arthropods (Comer et al., 1990; Pérez De León et al., 2006), including Simulium vittatum blackflies, which can transmit vesicular stomatitis New Jersey virus to cattle under experimental conditions (Mead et al., 2009), and Phlebotomus argentipes sandflies, which efficiently transmitted Chandipura virus to laboratory mice (Mavale et al., 2007). It is currently unknown whether MERDV has the capacity to replicate in vertebrate hosts, but should its host range include vertebrates, it is unlikely that virus persistence in nature is dependent upon the bite of haematophagous arthropods. Male mosquitoes do not feed on blood, and thus a significant bias between male and female mosquito MIRs would be expected if that would be the case. Instead, the similar MIRs that we determined for MERDV are more compatible with vertical and venereal transmission. The occurrence of rhabdovirus transmission through these modes, in addition to horizontal transmission, has been demonstrated, for example, in Aedes aegypti mosquitoes for Chandipura virus (Mavale et al., 2007).

### Table 4. Comparison of MIRs for MERDV in female and male C. quinquefasciatus in the Yucatan Peninsula of Mexico, 2007–2008

<table>
<thead>
<tr>
<th>Study site</th>
<th>No. mosquitoes tested*</th>
<th>No. pools tested (positive)</th>
<th>MIR†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
<td>Male</td>
<td>Total</td>
</tr>
<tr>
<td>Merida‡</td>
<td>3018</td>
<td>3145</td>
<td>6163</td>
</tr>
<tr>
<td>Tixkokob‡</td>
<td>5020</td>
<td>4051</td>
<td>9071</td>
</tr>
<tr>
<td>Total</td>
<td>8038</td>
<td>7196</td>
<td>15 234</td>
</tr>
</tbody>
</table>

*All mosquitoes were female.
†Expressed as the number of positive mosquito pools per 1000 mosquitoes tested.
‡Mosquitoes were collected outdoors using mosquito magnets.

### Table 5. MIRs for MERDV in selected Aedes, Anopheles, Ochlerotatus and Psorophora spp. mosquitoes

<table>
<thead>
<tr>
<th>Species</th>
<th>No. mosquitoes tested*</th>
<th>No. pools tested (positive)</th>
<th>MIR†</th>
<th>Study site where positive pools were collected‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aedes aegypti</td>
<td>419</td>
<td>20 (0)</td>
<td>0</td>
<td>Tixkokob, Merida, Cozumel Island, Sian Ka’an and Tzucacab</td>
</tr>
<tr>
<td>Anopheles albimanus</td>
<td>727</td>
<td>20 (0)</td>
<td>0</td>
<td>Tixkokob, Merida, Cozumel Island, Sian Ka’an and Tzucacab</td>
</tr>
<tr>
<td>Anopheles crucians</td>
<td>691</td>
<td>20 (0)</td>
<td>0</td>
<td>Tixkokob, Merida, Cozumel Island, Sian Ka’an and Tzucacab</td>
</tr>
<tr>
<td>Anopheles vestitipennis</td>
<td>913</td>
<td>20 (0)</td>
<td>0</td>
<td>Tixkokob, Merida, Cozumel Island, Sian Ka’an and Tzucacab</td>
</tr>
<tr>
<td>Ochlerotatus taeniorhynchus</td>
<td>1000</td>
<td>20 (3)</td>
<td>3.0</td>
<td>Cozumel Island</td>
</tr>
<tr>
<td>Ochlerotatus trivittatus</td>
<td>734</td>
<td>20 (3)</td>
<td>4.1</td>
<td>Cozumel Island, Merida and Tzucacab</td>
</tr>
<tr>
<td>Psorophora cyanescens</td>
<td>1000</td>
<td>20 (0)</td>
<td>0</td>
<td>Tixkokob, Merida, Cozumel Island, Sian Ka’an and Tzucacab</td>
</tr>
<tr>
<td>Total</td>
<td>5484</td>
<td>140 (6)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*All mosquitoes were female.
†Expressed as the number of positive mosquito pools per 1000 mosquitoes tested.
‡Mosquitoes were collected using mosquito magnets at five study sites (Tixkokob, Merida, Cozumel Island, Sian Ka’an and Tzucacab) in 2007–2008.
Discovery of a novel rhabdovirus in Mexican mosquitoes

Fig. 3. Relative UpA and CpG frequencies in the L protein ORF of different rhabdovirus species. UpA and CpG frequencies were calculated in two different ways. (a) In each sequence, the numbers of UpA and CpG dinucleotides, and A, C, G and U mononucleotides, were counted. Dinucleotide frequencies, $f_{XpY}$, were expressed relative to their expected frequencies, $f_X f_Y$, in the absence of selection. (b) To factor-out codon and amino acid usage, 1000 shuffled ORF sequences were generated for each virus sequence. In each shuffled sequence, the original amino acid sequence and the original total numbers of each of the 61 codons were maintained, but synonymous codons were randomly shuffled between the different sites where the corresponding amino acid is used in the original sequence. Next, the UpA and CpG frequencies in the original sequence were expressed relative to their mean frequencies in the codon-shuffled sequences. As codon usage is factored-out, the UpA and CpG relative frequencies tend to be less extreme in (b) compared with (a). Each point represents a single rhabdovirus sequence. Solid points correspond to species within defined genera, colour coded by genus (see key). Annotated open circles correspond to species that are currently unassigned at genus level, colour coded by host (or presumed host) taxa. Asterisks in the key indicate clades with uncertain host taxa: viruses in the unclassified ‘arthropod-infecting’ clades (yellow open circles) were isolated from arthropods but not from vertebrates; the sole representative of the genus *Tupavirus* was isolated from mammals but not from arthropods, although its phylogenetic position suggests that it may be arthropod-borne; the presence of viruses derived from vertebrates and viruses derived from arthropods in each of the unclassified ‘vertebrate (arthropod-borne)’ clades (brick-red open circles) suggests that all of these viruses are likely arboviruses. GenBank accession numbers of sequences used: NC_000855, NC_000903, NC_001542, NC_001560, NC_001615, NC_001652, NC_002251, NC_002526, NC_002803, NC_003243, NC_003746, NC_005093, NC_005974, NC_005975, NC_006429, NC_006942, NC_007020, NC_007642, NC_008514, NC_008527, NC_009527, NC_011532, NC_011542, NC_011639, NC_013135, NC_013955, NC_014136, NC_017714, NC_018381, NC_020803, NC_020804, NC_020805, NC_020806, NC_020807, NC_020808, NC_020809, NC_020810, NC_022580, NC_022581, NC_022756, NC_025251, NC_025253, NC_025255, NC_025341, NC_025342, NC_025353, NC_025354, NC_025355, NC_025356, NC_025357, NC_025358, NC_025359, NC_025360, NC_025361, NC_025362, NC_025363, NC_025364, NC_025365, NC_025366, NC_025371, NC_025376, NC_025377, NC_025378, NC_025379, NC_025380, NC_025382, NC_025384, NC_025385, NC_025387, NC_025389, NC_025391, NC_025392, NC_025393, NC_025394, NC_025395, NC_025396, NC_025397, NC_025398, NC_025399, NC_025400, NC_025401, NC_025405, NC_025406, NC_025408.
et al., 2005) and in phlebotomine sandflies for vesicular stomatitis Indiana virus (Tesh et al., 1972). A hallmark in the life cycle of insect-specific sigmaviruses is the exclusively vertical transmission through eggs and sperm (Longdon et al., 2011). Currently, no data are available to decide to what extent horizontal or vertical transmission contribute to the maintenance of MERDV in nature. Attempts to isolate MERDV by inoculation of Vero cells were unsuccessful, and together with the determined CpG and UpA dinucleotide usage biases, suggest that MERDV is likely not to infect vertebrates in nature. However, attempts to isolate MERDV in C6/36 cells also were unsuccessful thus far.

The inability to recover an isolate of MERDV in C6/36 cells is unexpected because this cell line supports the replication of a diverse range of mosquito-associated viruses, including several rhabdoviruses recently identified in Culex and Ochlerotatus spp. mosquitoes (Quan et al., 2010; Kuwata et al., 2011; Coffey et al., 2014; Vasilakis et al., 2014). However, it is not without precedent; Manitoba rhabdovirus from Culex tarsalis is reported to not propagate in C6/36 cells, although it does replicate in Vero, primary chick embryo and mouse neuroblastoma cells (Artso, 1991). Shortcomings in sample handling and possible failures in the cold-chain during transport are unlikely for the failure to isolate MERDV because one-quarter of our mosquito homogenates induced virus-like cytopathic effect when inoculated onto C6/36 cells, indicating that other, undetermined viruses did successfully propagate. Another possibility is that MERDV does not actively replicate in mosquitoes. It cannot be excluded that some of the field-collected mosquitoes passively carried MERDV without the virus being capable of replicating in the mosquito, despite the detection over multiple years, several locations and specific species. Alternatively, we may have discovered another example of endogenous viral elements analogous to the rabdovirus-like sequences described previously in various insect hosts (Nene et al., 2007; Katzourakis & Gifford, 2010; Li et al., 2015). However, such endogenous viral elements have thus far been reported to concern only partial sequences, at best covering one gene, but never what appears to be a complete, functional genome as we have found for MERDV.

Amino acid sequence alignments and phylogenetic analyses indicated CTRV as the closest known relative of MERDV. CTRV was isolated in C6/36 cells from C. tritaeniorhynchus in Japan (Kuwata et al., 2011), and later detected in Culex, Aedes, Anopheles and Armigeres spp. mosquitoes in China (Li et al., 2015; Shi et al., 2015). Recent studies have also shown that NIID-CTR cells, which were established from C. tritaeniorhynchus embryos, are persistently infected with CTRV (Gillich et al., 2015). CTRV establishes a non-cytoplastic infection and, similar to sigmaviruses, employs vertical transmission. However, in contrast to sigmaviruses, CTRV replicates in the nucleus of the infected cell similar to nucleorhabdoviruses and it is the only known rhabdovirus that requires the cellular splicing machinery for its mRNA maturation. The coding region for the L protein of CTRV is interrupted by a 76 nt intron (Kuwata et al., 2011). Inspection of the MERDV sequence provided no evidence for the use of splicing similar to CTRV, suggesting that MERDV may not require a nuclear phase. Our proposed transcriptional signals match those confirmed for CTRV. Whilst conservation of the termination signal sequence is seen in comparison to other rhabdoviruses, including the Drosophila-specific sigmaviruses and Moussa virus, a potentially mosquito-specific rhabdovirus from Culex decans mosquitoes (Quan et al., 2010), the initiation signal sequence differs from the two other viruses. Additionally, whereas sigmaviruses and Moussa virus do not show overlap of genes, the G and L genes of CTRV and both the M/G and G/L genes of MERDV show overlap. Other rhabdoviruses also possess overlapping transcription termination and transcription initiation sequences in their genomes including two more recently discovered mosquito-associated rhabdoviruses: Malpais Spring virus and Oak Vale virus (Quan et al., 2011; Vasilakis et al., 2013). Indeed, the positioning of the initiation signal of the downstream gene in front of the termination signal of the preceding gene or the use of splicing are not unprecedented in mononegaviruses, e.g. in human metapneumovirus or in bornaviruses, where these mechanisms have been hypothesized to adjust transcription levels possibly in conjunction with persistent infection, or attenuate gene expression in addition to the 3′→5′ transcriptional gradient characteristic for mononegaviruses (Collins et al., 1987; Schneemann et al., 1994; Schneider et al., 1994).

The translated primary sequences of MERDV N, G and L ORFs show significant homology to the respective ORFs of other rhabdoviruses. However, as also observed for CTRV, sigmaviruses and Moussa viruses, the P and M ORFs are far more diverse and do not exhibit homology to any other sequences in GenBank or to each other (except for the M proteins of sigmaviruses, which have a very distant relationship to the corresponding protein of Flanders virus). In addition, sigmaviruses contain an additional ORF (designated ORF X) between the P and G ORFs.

Although CTRV is the closest known relative of MERDV, these two viruses exhibit considerable sequence dissimilarity. The L proteins of MERDV and CTRV, which represent the most conserved genome regions of the rhabdoviruses, show an amino acid divergence of 56 %. This is reaching the divergence observed between rhabdoviruses belonging to different genera, which is commonly in a range of 47–83 % (Table S1, available in the online Supplementary Material). As species and genus demarcations for rhabdoviruses also include factors such as biological characteristics (e.g. host range) and serological cross-reactivity, additional work will be needed to accurately determine the taxonomic status of MERDV within the family Rhadloviridae.

In summary, we provide evidence that a novel rhabdovirus occurs in mosquitoes in the Yucatan Peninsula of Mexico.
This apparent virus, provisionally named MERDV, is most closely related to CTRV, although it shows considerable sequence and biological divergence. Our findings underscore the vast diversity of this virus family, highlighting the power of next-generation sequence technology in the discovery of novel viruses and provide the basis for improved surveillance programmes to gain better insights into arbovirus evolution.

**METHODS**

**Mosquito collections.** Mosquitoes were collected in five study areas in the Yucatan Peninsula of Mexico: Cozumel Island, Merida, Sian Ka’an, Tixkokob and Tzucacab. Descriptions of these study areas are provided elsewhere (Farfan-Ale et al., 2009, 2010). Collections were made in 2007 and 2008 using mosquito magnets (all five study areas) and in 2013 using CDC backpack-mounted aspirators (Merida only). Mosquito magnets Pro-Liberty (American Biophysics) were baited with propane and octenol, and placed outdoors. Mosquito magnets were turned on between 16:00 and 18:00 and collection nets were replaced the following morning between 06:00 and 09:00. CDC back-pack mounted aspirators were used to collect resting mosquitoes inside private residences. Mosquitoes were transported alive to the Universidad Autonoma de Yucatan, frozen at -80°C, and identified on chill tables according to species and sex using morphological characteristics (Darsie, 1996). Mosquitoes were transported on dry ice from the Universidad Autonoma de Yucatan to Iowa State University by World Courier.

**High-throughput sequencing.** Mosquitoes were homogenized as described previously (Farfan-Ale et al., 2009) and total RNA was extracted using TRIzol (Invitrogen) following the manufacturer’s instructions. Extracts were reverse transcribed using SuperScript III (Thermo Fisher) with random hexamers. The cDNA was RNase-H-treated prior to second-strand synthesis with Klenow fragment (NEB). The generated double-stranded cDNA was sheared to a mean fragment size of 200 bp using a Covaris focused-ultrasonicator E210 and the manufacturer’s standard settings. Sheared products were purified (Agencourt Ampure DNA purification beads; Beckman Coulter) and libraries constructed. Sheared nucleic acid was ended-repaired, DNA-tailed, ligated to sequencing adapters (NEBNext modules; NEB), PCR-amplified (Phusion High-Fidelity DNA polymerase; NEB) and quantified by an Agilent Bioanalyzer for sequencing. Sequencing on the Illumina HiSeq 2500 platform (Illumina) resulted in a mean of 180 million reads per lane. Samples were de-multiplexed using Illumina software and FastQ files generated. Data were quality-filtered and trimmed (Slim-Filter), and identified using Illumina software and FastQ files generated. Data were quality-filtered and trimmed (Slim-Filter), and identified using Illumina software and FastQ files generated. Data were quality-filtered and trimmed (Slim-Filter), and identified using Illumina software and FastQ files generated. Data were quality-filtered and trimmed (Slim-Filter), and identified using Illumina software and FastQ files generated.

**RT-PCR and Sanger sequencing.** Total RNA was analysed by RT-PCR using MERDV-specific primers RHAB-for (5’-CTACGACGTTAGCTTAGCTGCG-3’) and RHAB-rev (5’-CAGATTGGCGTGGGGTC-3’), which target a 490 nt region of the L protein gene, or RHAB-121-for (5’-AAGGCCGACTGATGATCG-3’) and RHAB-121-rev (5’-GACATCGAAGAGATGTTATCTG-3’), which target a 220 nt region of the N protein. cDNAs were generated using SuperScript III reverse transcriptase (Invitrogen), and PCRs were performed using Taq polymerase (Invitrogen) and the following cycling conditions: 94°C for 3 min, then 35 cycles of 94°C for 30 s, 56°C for 45 s and 72°C for 1 min followed by a final extension at 72°C for 8 min. RT-PCR products were purified using the Purelink Gel Extraction kit (Invitrogen). Sanger sequencing was performed using a 3730x1 DNA sequencer (Applied Biosystems).

**Amino acid sequence alignments and prediction algorithms.** The predicted amino acid sequences of MERDV were compared to all other sequences in GenBank by application of TBLASTN (Altschul et al., 1990). Per cent amino acid identities and similarities of select rhabdovirus protein sequences were calculated using CLUSTAL W 2.1 (http://www.ebi.ac.uk/clustalw2). The following prediction algorithms were used for the amino acid sequence analysis: NetNGlyc 1.0 server (for the identification of potential N-linked glycosylation sites), SignalP 4.1 server (for the identification of potential signal peptides), TMHMM server 2.0 (for the identification of potential transmembrane domains and cytoplasmic domains) and NetPhos 2.0 server (for the identification of potential PKC and TYR phosphorylation sites).

**Virus isolation.** An aliquot (200 µl) of each supernatant that tested positive for MERDV RNA was added to 2 ml Liebowitz’s L15 medium (Invitrogen) supplemented with 2 % FBS, 2 mM L-glutamine, 100 U ml−1 penicillin, 100 µg ml−1 streptomycin and 2.5 µg fumonisin B1. Samples were filtered using a 0.22 µm filter and inoculated onto subconfluent monolayers of Aedes albopictus C6/36 cells in 75 cm² flasks. Cells were incubated for at least 1 h at room temperature on an orbital shaker. Another 12 ml L15 maintenance medium was added to each flask and cells were incubated at 28°C for 7 days. After two additional blind passages, supernatants were harvested and tested by RT-PCR for the presence of MERDV RNA.
ACKNOWLEDGEMENTS

The authors thank Valeria Bussetti for expert technical assistance. This study was supported by the National Institutes of Health (5R21AI067281, AI057158, 5R21AI067281 and AI088647), the US Department of Defense and an intramural grant from Iowa State University. A. E. F. is supported by a grant from the Wellcome Trust (106207).

REFERENCES


